DOCUMENT RESUME

ED 030 085 AL 001 726

By-Sampson, Geoffrey

Noun-Phrase Indexing, Pronouns, and the 'Definite Article".

Pub Date Feb 69

Note-13p.; Paper presented at the Annual Meeting of the Linguistic Society of America, New York, December 30, 1968.

EDRS Price MF-$0.25 HC-$0.75

Descriptors-*Linguistic Theory, *Nominals, *Phrase Structure, Pronouns, Semantics, *Transformation Generative Grammar

Identifiers-Definite Articles, Reflexive Pronouns

Chomsky has suggested that certain lexical items, which he calls "referential items," should be given integer markers (or "indices") and that the noun-phrase deletion transformation which creates reflexive pronouns should be limited to cases where the noun-phrase to be deleted is fully identical to the antecedent noun-phrase, including identity of index. As long as it is conceded that the notion "referential item" can be specified somehow, it seems clear that the idea of indexing, the essence of Chomsky's suggestion, is correct. However, the author objects to the proposal as it stands for two reasons: (1) The addition to the general theory of language seems disproportionately costly in relation to the problem it is introduced to solve, and (2) Chomsky leaves it quite unclear what status deep structures have when the two noun-phrases have the same index but different lexical items. This type will greatly outnumber the deep structures in which the two noun phrases are fully identical. The author suggests that the phenomena associated with noun-phrase indexing are an area of language where any generative grammar is going to be of limited explanatory power. He proposes, using a "mentalistic" approach, an ordered set of rules for producing an English noun phrase to refer to a given referent. (DO)

EDO 30085

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NOUN-PHRASE INDEXING, PRONOUNS, AND THE 'DEFINITE ARTICLE'

Geoffrey Sampson

February 1969

Linguistic Automation Project
YALE UNIVERSITY
New Haven, Connecticut

AL 001 726

NOUN-PHRASE INDEXING, PRONOUNS, AND THE 'DEFINITE ARTICLE'[1]

The device of indexing certain lexical items in deep structures with integers was first introduced into linguistic theory by Chomsky (1965: 145-146). Chomsky was led to this innovation by considering the difference in meaning between pairs of sentences such as 'the teacher hates himself' versus 'the teacher hates the teacher'. In 'the teacher hates himself', the reference of 'the teacher' must be the same as that of 'himself'; while in 'the teacher hates the teacher', on the other hand, the two tokens of the item 'the teacher' must have different reference. Within a generative theory of language, it seems desirable for the 'himself' in 'the teacher hates himself' to derive from an underlying 'the teacher', so that the semantic component can specify that the same reference is involved. But if the distinction in meaning between the two sentences is to be reflected in their 'semantic interpretations', as it surely must be, their deep structures must differ in some respect.

Chomsky's suggestion is that certain lexical items, which he calls 'referential items', should be given integer markers, or indices as I would prefer to call them; and that the noun-phrase deletion transformation which creates reflexive pronouns should be limited to cases where the noun-phrase to be deleted is fully identical to the antecedent noun-phrase, including identity of index. So 'the teacher hates himself' would derive from 'the teacher$_j$ hates the teacher$_j$' -- I use subscripted j, k, and so on to stand for arbitrary indices -- while 'the teacher hates the teacher' would derive from 'the teacher$_j$ hates the teacher$_k$'.

Chomsky admits that it is not obvious what is to be counted as a 'referential item'.

As long as we concede that the notion 'referential item' can be specified somehow, it seems clear that the idea of indexing, the essence of Chomsky's suggestion, is correct. But there are two objections to be made to the proposal as it stands.

In the first place, the addition to the general theory of language seems disproportionately costly in relation to the problem it is introduced to solve. For it is fairly obvious that the indexing device cannot be restricted to cases where it is needed to distinguish identical from, as it were, homophonous noun-phrases. If 'the teacher' is a referential item in 'the teacher hates himself' and 'the teacher hates the teacher', then it must equally be a referential item in 'the teacher sold the car', 'Mary liked the teacher', and infinitely many other sentences where no possibility of ambiguity arises. In fact, almost every noun-phrase must be indexed.

And it is not just a case of adding an extra rule of a familiar type. The new rule is of a quite special type: instead of operating with a finite set of features established for the language, it involves an unlimited range of integers.

So all in all this has to be counted as a major addition to linguistic theory. But the only problem which Chomsky mentions as motivating it is the phenomenon of reflexivisation.

The second objection to the proposal, which to my mind is of greater weight, concerns its arbitrariness. Suppose that, in some given situation, we know that 'the teacher' is 'my landlord'; that is, that the reference of the phrase 'the teacher' is the same as that of the phrase 'my landlord'.

Then surely it is quite arbitrary, so far as the semantic component is concerned, to say that 'the teacher hates himself' derives from 'the teacher$_j$ hates the teacher$_j$', rather than from 'the teacher$_j$ hates my landlord$_j$'? But TG theory is forced to allow only the first on these derivations; the derivation from 'the teacher$_j$ hates my landlord$_j$' has to be ruled out because it would violate the principle of recoverability of deleted elements. In fact, Chomsky leaves it quite unclear what status deep structures of this latter type have, where the two noun-phrases have the same index but different lexical items; but of course they will greatly outnumber the deep structures in which the two noun-phrases are fully identical.

It is fairly clear that indices are logically partners of noun-phrases as wholes, rather than of individual lexical items. Chomsky (1965) suggests that the index is associated just with the head noun of the phrase, but this is not essential to his exposition. That the index belongs to the noun-phrase as a whole explains the fact that reflexivisation deletes the whole noun-phrase. Another fact that would be hard to explain otherwise is that noun-phrases with no head noun may be reflexivised: for instance 'the younger one admires himself', where the word 'one' is certainly only a place-holding element supplied in the surface structure, rather than a meaningful element of the deep structure.

Now if we regard a noun-phrase as a pairing of an index with a realisation in lexical items, I claim that from a semantic point of view the index has a logical priority over the lexical items. As I have tried to suggest with my examples 'the teacher$_j$ hates the teacher$_j$' and 'the teacher$_j$ hates my landlord$_j$', the most satisfying account will be one in which identity or difference of noun-phrases will be relevant only with

respect to indices, and not with respect to lexical items.

In this paper I want to suggest that the phenomena associated with noun-phrase indexing are an area of language where any generative grammar is going to be of limited explanatory power. By a 'generative' grammar, I mean one which aims at producing by algorithm all and only the sentences of a language; a sentence being a pairing of a semantic content and a phonetic expression. This kind of grammar is obviously logically quite distinct from what I would call a 'communicational' grammar; that is, a grammar which offers an algorithm for deriving a phonetic expression from a given semantic content, and a semantic content from a given phonetic expression. Fodor and Garrett (1966) have pointed out that communicational linguistics, as I call it, is a reasonable subject for study; and the main efforts toward a communicational theory at the moment, to my knowledge, are Sydney Lamb's'stratificational theory' (see e.g. Lamb, 1966) and Wallace Chafe's 'generative semantics' (see e.g. Chafe, 1967).

The account of noun-phrases I suggest assumes a mentalist approach to language; this shouldn't be objectionable to linguistics who accept Chomsky (1959) and Katz (1964). Independently of linguistics, a framework very similar to the one I propose is advocated by the logician Peter Geach of Birmingham University (Geach, 1957).

I suggest that the mind of a person entering into a speech situation should be regarded as containing a number of entities which I shall call 'referents'; for convenience, we may think of a referent as labelled by an integer index. Associated with each referent will be information sufficient to determine how the referent could appropriately be described

using general terms -- common nouns, adjectives and verbs. For instance, 'j' will be a referent associated with the information that it is a teacher, and that it is my landlord.

Also, some, but not all, of the referents will be labelled by individual terms or proper names; names appear to be the equivalent, within a language, of the indices we are positing in our theory about language. Thus, 'j' may be indexed with the name 'Mr. Smith'.

Obviously, to a large extent the set of referents will correspond to physical features of the person's environment; but it is clear that that is not the whole story, because we can talk about unicorns just as easily,and formally in exactly the same way, as we talk about taxicabs.

Now a sentence can be thought of as an operation on the hearer's set of referents; and let me make the simplifying assumption that the speaker has a perfect knowledge of the hearer's state of mind in this respect. Then we may distinguish two types of noun-phrase. In a sentence such as 'I bought a red car yesterday', the noun-phrase 'a red car' is an instruction to the hearer to add a new referent to his referent stock. On the other hand, in 'I bought the red car yesterday', the noun-phrase 'the red car' constitutes an attempt to single out a particular one of the referents already in the hearer's referent-stock. I shall call these types of noun-phrase 'establishing' and 're-identifying' respectively.

The distinction between 'establishing' and 're-identifying' noun-phrases is obviously related to the commonly-drawn distinction between 'indefinite' and 'definite'. Certainly, noun-phrases beginning with the indefinite article are always establishing, and those with the definite article are always re-identifying.[2] But some qualifications must be made.

In the first place, linguists who have discussed 'indefinite' and 'definite' as properties of noun-phrases (for instance Postal, 1966) have tended to present '±definite' as a semantic feature or attribute, on a logical par with, for instance '±animate'. An extreme position of this type is adopted by Carlota Smith (1964): she maintains that definiteness is actually a semantic gradient.

But the distinction between re-identifying and establishing is a simple all-or-nothing opposition, and one which serves a completely different function from semantic classifications like animate versus inanimate or male versus female.

Secondly, Postal maintains that all noun-phrases are either definite or indefinite. But a few noun-phrases, at least, are neither re-identifying nor establishing. Chafe (1968) has pointed out that while the object of the sentence 'I am looking for some elephants' (with unstressed some) is establishing (to use my own terminology), and that of 'I am looking for the elephants' is re-identifying, the object of 'I am looking for elephants' has to be regarded as neither of these; rather, it is a modification of the verb, serving much the same function as an adverbial.

A further qualification is that re-identifying noun-phrases exhibit a distinction between restrictive and non-restrictive relatives. Chomsky points out (1965:217) that non-restrictive relatives are complements of the full noun-phrase. As I see it, Chomsky's term 'referential item' covers the set of establishing noun-phrases, together with re-identifying noun-phrases minus any non-restrictive relatives. I shall call all these 'noun-phrases proper', or NPPs for short.

At this point I should like to leave aside the establishing NPPs, and discuss the rules the speaker uses to construct a re-identifying NPP which will do the job of unambiguously singling out one of the

hearer's referents. Most re-identifying NPPs in English begin with the definite article the. It might at first seem that if an NPP like 'the red car' is to be used correctly, it must be the case that the hearer has only a single referent which could be described as a red car.

But that is really not true. It is easy to imagine a situation where both speaker and hearer are aware of the existence of several particular red cars; yet the phrase 'the red car' will be understood as referring unambiguously to one among them which is closest to the focus of attention. So we must consider the hearer's stock of referents to be arranged in some kind of mental space at different distances from this focus.

Now we can write an ordered set of rules for producing an English NPP to refer to a given referent. First, if the referent has already occurred in the kernel sentence, the NPP takes the feature 'reflexive'. This will account for the usage of the reflexive pronoun in 'the teacher hates himself', and also the emphatic usage in, for instance, 'the landlord himself constructed the building', which becomes a special case in English of the common phenomenon of reduplication for emphasis.

Whether or not this condition is met, one then goes on to the next rule. The rest of the rules are disjunctively ordered -- if one of them is satisfied, the NPP is complete. The second rule is: if the referent is the speaker or hearer, the first or second person pronoun is used. Third, if the referent is the nearest one to the focus of attention, the appropriate third person pronoun is used: he, she, or it.

Fourth, if the referent has a name, the NPP consists of just that name. At one time I thought the rule for using names was an optional rule; but it seems to me now that this rule is as obligatory as all the others. That is, if we know that 'the teacher' is 'Mr. Smith', it is

incorrect to refer to him as 'the teacher' -- just as it is incorrect to refer to oneself or to one's hearer by name instead of using the appropriate pronoun.

The only exceptions I can think of to the name rule all fall under the heading of 'taboo names'. If a referent is a person of high status, it may be impolite to use his name: thus, say, a senior civil servant will say 'the Minister' rather than 'Mr. Brown'. But this isn't a contradiction of the basic rule, just a modification of it.

Fifth: if none of the previous rules applies, the NPP consists of the word the, followed by enough lexical items to rule out all the other referents as near or nearer to the focus of attention than the target referent. For instance, if one refers to a referent, say k, by means of the NPP 'the red car', it must be the case that k can be described as a red car, and no other referent, say l, which could also be described as a red car is as near to the hearer's focus of attention as k. Of course, in English the word the will be deleted before possessives.

There is nothing new in the condition for reflexivisation which I give in rule 1. But notice that within the communicational framework, the question whether 'the teacher hates himself' derives from 'the teacher$_j$ hates the teacher$_j$', or from 'the teacher$_j$ hates my landlord$_j$', becomes a pseudo-problem. The true origin of the sentence is 'j hates j'; the fact that the first token of 'j' happens to be realised as 'the teacher' rather than as 'my landlord' is irrelevant to the second token, which has to be realised as the reflexive pronoun anyway. The question of assigning this second index 'j' a noun realisation simply never arises.

The account of noun-phrases I give raises a considerable problem about the nature of the mental space, and how the position of the referents

and the focus is determined. But I believe this is a genuine problem which would sooner or later have to be faced by any theory of language, rather than a pseudo-problem created by the particular framework I advocate.

Most of the work on this subject I am aware of is by Charles Taber of the Hartford Seminary Foundation: in a 1966 paper read to the Yale Linguistics Club, he has pointed out that, apart from recency of previous mention in discourse, continuity of syntactic rôle is a relevant dimension. Taber observes that in the sentence ' A man saw a little boy crying, so he asked him: "What's the matter?"', we have no difficulty in identifying he with a man and him with a little boy, because the rôles of subject and object remain unchanged.

Again, Archibald Hill (1966:228-9) has pointed out that physical proximity of the referent in the outside world can be relevant: thus, in the sentence 'Please open the window', other things being equal 'the window' will refer to the nearest window.

The communicational approach to noun-phrases raises a further problem, which I can only hint at here. This concerns Frege's distinction between meaning and reference, or intension and extension. The validity of the distinction is usually demonstrated by giving pairs of phrases -- we can continue to use 'the teacher' and 'my landlord' -- which differ in meaning but have the same reference. Quine (1953:Chapter 3) maintains that the job of semantics is to explicate meaning, as opposed to reference.

Now the validity of the argument depends on our agreeing that the reference of the phrase 'the teacher' is a physical object, in this case a human being. But surely the reference of a noun-phrase must be taken

to be the mental referent within the hearer, and not the physical entity in the outside world which that referent may be associated with. If this weren't so, it's hard to see how consistent discourse would be possible about non-existent objects, for instance characters in a novel: because the referent of the name of each character would be the same, namely the null set.

But now, if the reference of a noun-phrase is a mental entity, it is no longer obvious that the reference of 'the teacher' is identical to the reference of 'my landlord'. In fact, in normal usage the occurrence of a sentence like 'the teacher is my landlord' implies that the two noun-phrases have different referents for the hearer; otherwise the utterance would be pointless. This opens up a very wide area of speculation indeed, which I must leave untouched here.

Footnotes

[1] This work was read as a paper to the Annual Meeting of the Linguistic Society of America, New York City, 30 December 1968. No substantial revisions have been made. The work was supported by a grant from the National Science Foundation.

[2] Lauri Karttunen (personal communication) has pointed out that this generalisation is valid for the indefinite article only in declarative, non-negative, non-conditional clauses, and for the definite article only in the absence of certain quantifiers.

Bibliography

Chafe, W.L. (1967) Language as symbolisation. Language 43, 57-91.

Chafe, W.L. (1968) English noun inflection and related matters from a generative semantic point of view. Project on Linguistic Analysis Reports, Second Series, no.6. Department of Linguistics, University of California at Berkeley.

Chomsky, Noam (1959) A Review of B.F. Skinner's Verbal Behavior. Language 35, 26-58.

Chomsky, Noam (1965) Aspects of the Theory of Syntax, Cambridge, Mass.: M.I.T. Press.

Dinneen, F.P. (S.J.), ed. (1966) 17th Annual Round Table Meeting on Linguistics and Language Studies. Washington, D.C.: Georgetown University Press.

Fodor,J. and Garrett, M. (1966) Some reflections on competence and performance.In J.Lyons and R.J. Wales, ed.: Psycholinguistic Papers. Edinburgh: Edinburgh University Press.

Geach, Peter (1957) Mental acts. London: Routledge and Kegan Paul.

Hill, A.A. (1966) A re-examination of the English articles. In Dinneen, ed., 1966: 217-231.

Katz, J.J.(1964) Mentalism in linguistics. Language 40, 124-137.

Postal, P.M. (1966) On so-called 'pronouns' in English. In Dinneen, ed. 1966: 177-206.

Quine, W. van O. (1953) From a logical point of view. Cambridge,Mass.: Harvard University Press.

Smith, C.S. (1964) Determiners and relative clauses in a generative grammar of English. Language 40, 37-52.